

What is claimed is:

1           1.     A method for use in a parallel database system having plural nodes,  
2 comprising:  
3                 providing random number generators in the plural nodes;  
4                 generating, in parallel, random numbers using the random number  
5 generators in the plural nodes; and  
6                 performing random sampling using the generated random numbers.

1           2.     The method of claim 1, wherein providing the random number generators  
2 in the plural nodes comprises providing random number generators in at least some of the  
3 plural nodes.

1           3.     The method of claim 1, wherein the random number generators in the  
2 plural nodes are part of a first set of random number generators, the method further  
3 comprising using at least one other random number generator to generate random  
4 numbers provided as seeds to the first set of random number generators.

1           4.     The method of claim 1, further comprising determining a number of  
2 random samples to select in each node based on the generated random numbers in the  
3 node.

1           5.     The method of claim 1, further comprising:  
2                 providing a plurality of parameters;  
3                 adjusting values of the parameters based on values of the random  
4 numbers; and  
5                 determining a number of random samples to select based on the  
6 parameters.

1           6.     The method of claim 5, wherein providing the parameters, adjusting the  
2 values of the parameters, and determining the number of random samples are performed  
3 in each of the plural nodes.

1           7.     The method of claim 5, further comprising:  
2                 associating the parameters with corresponding predefined ranges;  
3                 determining which range each random number falls within; and  
4                 adjusting the value of one of the parameters based on the determined range  
5 of each random number.

1           8.     The method of claim 7, wherein adjusting the value of the one parameter  
2 comprises incrementing the value of the one parameter.

1           9.     The method of claim 8, wherein incrementing the value of the one  
2 parameter occurs in response to each occurrence of a random number in the determined  
3 range.

1           10.    The method of claim 7, further comprising communicating certain of the  
2 parameters between nodes, each node determining the number of random samples  
3 based on the communicated parameters.

1           11.    The method of claim 1, further comprising:  
2                 defining plural ranges; and  
3                 counting a number of occurrences of random numbers in each of the plural  
4 ranges.

1           12.    The method of claim 11, further comprising:  
2                 storing plural parameters having values set to represent the number of  
3 occurrences of random numbers in corresponding ranges.

1           13.    The method of claim 1, further comprising:  
2                 generating random number seeds at one node;  
3                 sending the random number seeds to plural nodes from the one node; and

4 the random number generators in the plural nodes using corresponding  
5 random number seeds.

1 14. A database system comprising:  
2 a plurality of nodes, wherein each of at least two of the plurality of nodes  
3 comprises:  
4 a storage for storing tuples of a relation; and  
5 a controller adapted to generate random numbers,  
6 the controller adapted to determine a number of random samples to  
7 generate in the node using the random numbers.

1 15. The database system of claim 14, wherein each node further comprises a  
2 processor, and wherein the controller is a software program executable by the processor.

1 16. The database system of claim 14, wherein one of the nodes comprises a  
2 first random number generator, and wherein each of the nodes comprises a second  
3 random number generator to generate the random numbers, the first random number  
4 generator to generate random number seeds for use by the second random number  
5 generators.

1 17. The database system of claim 16, wherein the one node communicates the  
2 random number seeds to corresponding nodes.

1 18. The database system of claim 16, wherein the plural nodes comprise nodes  
2  $i, i = 1 - L$ , wherein the first random number generator is adapted to generate random  
3 number seeds  $s_i, i = 1 - L$ , and wherein the one node is adapted to send each random  
4 number seed  $s_i$  to node  $i$ .

1 19. The database system of claim 18, wherein the second random number  
2 generator in each node  $i$  is adapted to generate random numbers  $r_1 \dots r_{M_i}$ , where  $M_i$   
3 represents a number of random numbers to be generated in node  $i$ .

1           20.    The database system of claim 19, the storage in each node  $i$  to store array  
2 elements  $A_{i1} \dots A_{iL}$ , the controller adapted to adjust the value of  $A_{ij}$ ,  $j$  equal to a value  
3 between 1 and  $L$ , based on which of plural predefined ranges each random number  $r$  falls  
4 within.

1           21.    The database system of claim 20, wherein each node is adapted to  
2 communicate certain of these array elements to other nodes.

1           22.    The database system of claim 21, wherein the controller in each node is  
2 adapted to sum values of the array elements to derive the number of random samples to  
3 generate.

1           23.    An article comprising at least one storage medium storing instructions that  
2 when executed cause a database system to:  
3                   generate random numbers in each of plural nodes of the database system;  
4 and  
5                   use the random numbers to determine a number of random samples to  
6 generate in each node.

1           24.    The article of claim 23, wherein the instructions when executed cause the  
2 database system to provide a first random number generator in each node to generate the  
3 random numbers.

1           25.    The article of claim 24, wherein the instructions when executed cause the  
2 database system to provide a second random number generator in one of the nodes to  
3 generate random number seeds for use by the first random number generators.

1           26.    The article of claim 25, wherein the instructions when executed cause the  
2 database system to distribute the random number seeds to the plural nodes.

1           27.    The article of claim 23, wherein the instructions when executed cause the  
2 database system to further:  
3                define parameters in each node; and  
4                adjust the parameters based on values of the random numbers in each  
5 node, wherein determining the number of random samples is based on the parameters.

1           28.    The article of claim 27, wherein the instructions when executed cause the  
2 database system to distribute certain of the parameters from each of the nodes to other  
3 nodes.

1           29.    The article of claim 28, wherein the instructions when executed cause the  
2 database system to sum the parameters at each of the nodes to derive the number of  
3 random samples.

1           30.    An article comprising at least one storage medium storing instructions  
2 executable in a database system having plural nodes, the instructions when executed  
3 causing a system to:  
4                generate random number seeds;  
5                communicate the random number seeds to the plural nodes; and  
6                generate random numbers in each node using at least one of the random  
7 number seeds.

1           31.    The article of claim 30, wherein the instructions when executed cause the  
2 database system to determine a number of random samples to generate in each node  
3 based on the generated random numbers.